Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230390

Two inherent paradoxes of machine learning approach to credit scoring

Jingnan Cao

38-279 Nanbeng Rd. Mangshi, Yunnan, China. 678400

jcao21@ucsc.edu

Abstract. This paper reviews existing literature on academic and commercial progress in the machine learning(ML) approach to credit scoring. In doing so this paper proposes two inherent paradoxes to ML(machine learning) approach to credit scoring. The first is the problem involving the transparency of data range and internal algorithms. The second paradox is the problem of data volume and accuracy.

Keywords: machine learning, credit scoring, bias, risks.

1. Introduction

An important instrument in credit activity is credit scoring. Traditionally, a fixed ratio linear formula is used to calculate the credit score of an individual or firm [1]. The most well-conceived system is FICO. The problem with traditional credit scoring is its reliance on customers' credit history and bank-certified financial information. For many so call "thin file" clients [2], finding such data for scoring can be difficult. In addition, traditional credit scoring requires an information-gathering procedure that is labor intensive, getting scored becomes costly for many small firms [3].

As an answer to the above shortcomings, the ML approach to credit scoring utilizes untraditional data such as customers' digital footprint and geological local, promising to expand credit accessibility. Traditional and ML approaches to credit scoring differ in two ways. First difference is the types of data being used. The second is how correlation is drawn between variable. ML approach to credit scoring relies on a multi-dimensional algorithm to draw complex correlation involving multiple variables. Contrasting to that simple algorithm of traditional method, this complexity introduces risks that we will discuss later.

Figure 1. FICO credit score composition.

DOI: 10.54254/2755-2721/6/20230390

2. Literature review

Existing literature on the subject of ML approach to credit scoring has two strings. One focuses on the exploration of possible data set that can be utilized in a credit scoring algorithm. Berg, et al. [4]'s study utilized customers' digital footprint on a German e-commerce website as bases for credit worthiness. In 2019, Matz, et al [5]. predicted customers' income level from Facebook status and likes alone. In 2020, Fu, et al. utilized customers' Wechat(popular Chinese social media) profile info such as profile pictures and profile names to predict creditworthiness, with moderate confidence. Wu, et al, [3] published a study in 2021 that utilizes a firm's association in the supply chain to predict the creditworthiness of a firm. These are studies with the most success.

The other string of research focuses on the social impact of ML approach to credit scoring, most comprehensive of which is from Hurley [2] Along with a group of lawyers, Hurley collected legal cases where customers have been negatively impacted by ML approach to credit scoring. While previous studies expressed concern with machine learning approach to credit scoring(or credit score in general), and some risk has been discussed extensively, none recognized the inherent paradox of machine learning approach to credit scoring. That is, in attempt to adapt for one risk, one inevitably enlarge or create another risk.This paper follows the second string of literature, attempt to illuminate on the second inherent paradoxes of ML approach to credit scoring.

3. Risk: transparency VS. score manipulation

In academic settings knowledge of algorithms is shared, however, in commercial environments, algorithms and their data collection are usually not public. Companies such as Zest finance hold their ML algorithms as company secret [2][6] In this section, we discuss a paradox regarding the transparency of ML algorithms in a commercial setting. There are inherent risks involved with the data collection and correlation of ML algorithms.

3.1. Risk with transparency

ZestFinance has been studied extensively by scholars on this subject. Scholars agree that ML algorithms with their data collection process are holden by companies as commercial secrets like ZestFinance [1][2][6]. The only access we have to this enigmatic process is companies' patents of the algorithms. This means for an average customer, it is impossible to know exactly how they get the credit score they have.

With the data collection process unknown to customers, customers have no way of knowing what contributes to their credit score, and thus have little to no bases for defending their creditworthiness. Also due to the multi-dimensional nature of ML algorithms, even when one input data is found the be false, customers cannot make a reasonable argument as to where the calculation of their credit score went wrong.

Hurley and a group of lawyers collected cases where individuals' credit score has been negatively affected by the utilization of ML algorithms by credit scoring agencies. One case they followed thoroughly is Kevin Johnson. In 2008, Kevin Johnson's credit limit has been lowered from $10800 to $3800 by American Express. The reason for lowering Kevin's score is perplexing. The company stated in the letter that Kevin has been considered a risky borrower since his credit card history show that Kevin shopped at places where customers have low repayment rate to American Express. When Kevin sought to explain, American Express was reluctant to give justification for their change of algorithm.

We can see two risks here. The first is that the data collection process was unauthorized by customers. Kevin and many others who fall victim to this new method of credit scoring did not anticipate their information would be collected and evaluated without their knowledge. In the traditional method of data collection, what types of data are collected were public knowledge, and credit card companies are authorized to provide the payment history of customers' credit cards, in particular, the frequency and amount of credit payments. Now with the "all data is credit data" approach, information can be collected from any traceable online or offline activities of customers.

DOI: 10.54254/2755-2721/6/20230390

Devices like website cookies silently collect customer data and sell them off to credit scoring agencies without customers' knowledge.

The second risk with Kevin's case is scoring by association. Kevin has been good with his finance management, spending a reasonable amount of credit each month, and paying back loans in time, all to build a good financial standing for himself[2]. If judging from individual bases, he is an exemplary case of a creditworthy individual. However since Kevin's new credit score was compiled by association rather than his action, his effort in improving his finance all went to waste. It could well be that Kevin shops at welfare stores because he comes from an unprivileged family, he is frugal so he can save up to improve his life. Since Kevin was born in an economically unprivileged community, his background puts a limitation on his finance. His ability to obtain credit was crucial to his financial improvement. Now with his lowered credit score, he is stuck in an unprivileged community. No amount of personal hard work could change that unless he chooses a "better" place to shop, which will further increase his financial pressure. From a societal standpoint credit scoring by association defeat, the purpose of credit as a tool for upward mobility.

3.2. Proxy for discrimination

To take this argument further, there are also concerns that ML algorithms for credit scoring will become bases for systematic discrimination. Using data on US mortgages, Fuster et al [7] find that African American and Hispanic borrowers are disproportionately less likely to benefit from ML algorithms utilization in credit scoring.

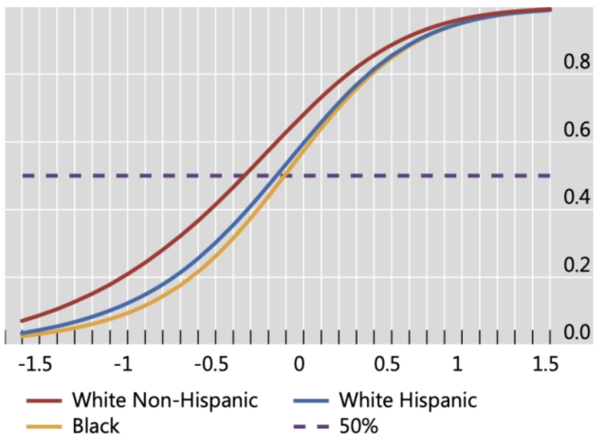

Figure 2. (Fuster et al): Algorithm bias by race.

Hurley[2] explains how ML algorithms can find proxies for discrimination such as race, disabilities nationality, and gender. From Kevin's example, we see one particular input is where the customer shops. In the United States, there are grocery shops that are dedicated to certain communities by race or nationality. Also by culture, different races will have different behavioral patterns. "Thirty percent of whites use their mobile phone as the sole internet connection compared to roughly forty-seven percent of Latinos and thirty-eight percent of blacks[8]". Though not explicitly an input variable, and possibly carefully avoided by algorithm developers, a factor of race, can be compiled from hundreds of racially biased variables that individually look harmless, but combined serve as a proxy for race.

This ability to proxy sensitive factors such as race is inherent to machine learning algorithms. To purge potential biases, one will have to abandon any biased data input which greatly limits the capability of ML algorithms; or manually adjust the weight of inputs by race, which creates more problems. Scoring by the association itself is greatly critiqued, it puts customers into categories by their economic background, strengthening the stratification of society by the economy.

DOI: 10.54254/2755-2721/6/20230390

3.3. Score manipulation

Som One solution to the transparency problem of ML algorithms is for firms to their algorithms and data collection process publicly available. However, this inevitably leads to the risk of score manipulation by customers. Credit score manipulation currently exists in a somewhat restrained form with the traditional method. Taking Kevin's case, for example, Kevin dedicated much of his energy to meeting the traditional algorithm – keeping a good credit history. There are various forms of credit score coach, just search any search engine "credit coach" and there will be guides on how to improve one's credit score.

In the traditional method, credit manipulation is done by keeping a good financial history. Though to some degree, this might affect the effectiveness of credit score. But keep in mind all the data collected by traditional methods are hard information, hard in the sense that all data are verifiable and non-manipulable [9] Since all inputs are directly linked to individuals' financial activities, individuals can only improve their scores by being responsible for their finance. The effort an individual is willing to give to improve their score only strengthens their case of responsibility.

ML algorithms utilize soft information that is not verifiable and thus manipulable [9]. Once customers know what information ML algorithms are considering as valuable inputs, customers can change their behavior patterns accordingly to seek improvement. Considering that many of the data points utilized by ML algorithms are data such as IP addresses, Facebook status, and Geological data that are easy to manipulate, improving one's score can be as simple as using a VPN to change the IP address to a rich neighborhood or frequently visiting websites associated with "good customers". Data manipulation of this sort doesn't require human attention, one can simply create programs to do these things automatically, and there is potential that these programs will be commercialized.

In 2019 Matz et al. [10] successfully developed an ML algorithm that predicts one's income level from Facebook status, like, and comments received alone. Along with this finding, a word map is compiled that associates the frequency of appearing words that have a strong association with both high and low-income individuals.

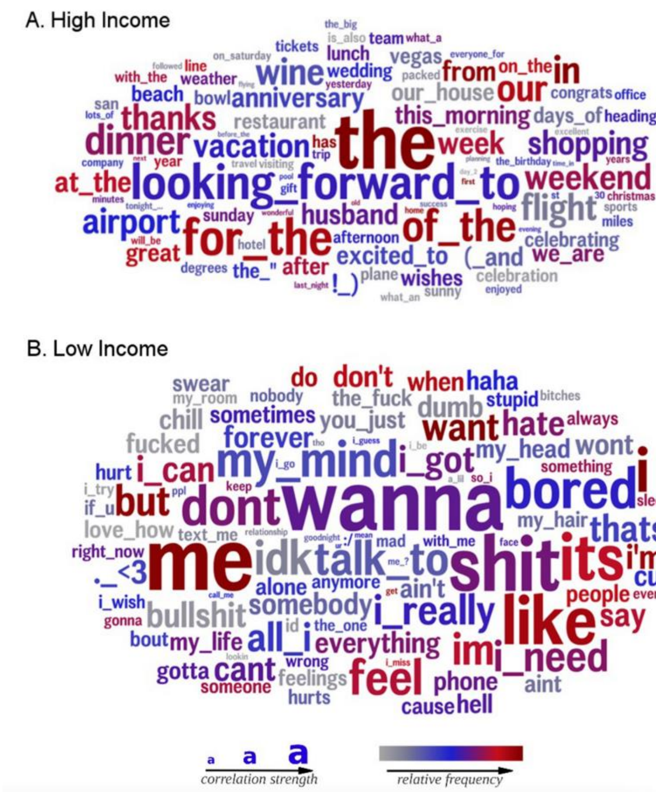

Figure 3. [10] Words in Status that have an association with low and high income.

It will be no surprise that these types of input will be utilized in credit scoring algorithms. Then, improving one's credit score comes down to changing one’s posting pattern, avoiding using words that indicate low income, and including words associated with high income in every social media post. Customer can go a step further to hire a “zombie” account that automatically likes and comment on a post.

No doubt that if a scoring agency's algorithms and data collection is made public, there will be a boom of these commercialized score manipulation tactics. Credit scoring by ML algorithms will then suffers in reliability. The risk not only affects borrowers but also customers who do not take advantage of such tactics.

4. Risk: inclusiveness VS. accuracy

The second paradox inherent in the ML approach to credit scoring is the balance of inclusiveness and accuracy. On the side of inclusiveness, the ML approach to credit scoring promise to solve the “thin profile” problem by including a wider range of data inputs. ML algorithms in general perform better the larger the given data set. At face value, the tremendous amount of information flowing into ML algorithms should increase the accuracy of the algorithms. However, such an advantage of ML algorithms is reliant on the assumption that the quality of data collected remains high quality. In practice, the data set suffers from the increasing volume.

To achieve the promised inclusiveness of the ML approach to credit scoring, ML algorithm developers may be sought to include a wider and wider range of data as their input variables. This is also a discernible trend in the academic adventure of developing newer ML algorithms for credit scoring. Some studies at the advent of such exploration yield very reliable results, [4] and the algorithms can perform with high confidence. Later studies that followed such as Fu’s adaptation using Wechat data are getting gradually weaker. The ML frameworks and logic of these studies are similar, but they mainly differ in the input values that are used. Along with the possible manipulation of ML algorithm input discussed in the prior section, a bigger range of data for the sake of inclusiveness may be at the cost of the accuracy of the algorithms.

Even assuming that the quality of data does not suffer, the reliability of correlations uncovered by ML algorithms may become trivial too, given a larger range of data. As Hurley quotes statistician Jame Cobioulus “One of the bedrock truths of statistics is that, given enough trials, almost any possible occurrence can happen . . . The more possible events that might take place, the more potential, albeit unlikely, ‘fluke’ events there are.” ML algorithms' correlation lacks an inherent reason, to begin with, all ML correlations are based solely on patterns among variables. It is possible that in pursuit of inclusiveness, ML algorithms will produce correlations that are contrary to the cause. As it is widely known in statistics, correlation is not equivalent to causality, correlation can just be a coincidence.

5. Conclusion

ML and its utilization in credit scoring are still in their infancy. There is still room for its improvement. What this paper discusses hopefully will serve as a suggestion in its implementation. Due to the scope of this paper, the content discussed here is admittedly simplified. And with the fast development of ML technology in recent years, some of the information may be outdated. The paper paints a somewhat grim view of ML's approach to credit scoring, but every piece of technology has its benefit and harm. ML utilization completes most of its promises, making credit scoring available for customers that are neglected by the traditional method. Given time, there will surely be clever ways to circumvent these two paradoxes of ML algorithms.

References

[1] Vidya CM. New Era of Credit Analysis(2018). Ushus-Journal of Business Management 2018, Vol. 17, No. 3, 39-46 ISSN 0975-3311| https://doi: 10.12725/ujbm.44.5

DOI: 10.54254/2755-2721/6/20230390

[2] Hurley, M., & Adebayo, J. Credit scoring in the era of big data(2016).. Yale JL & Tech., 18, 148-216.
[3] [Wu, Jing and Zhang, Zhaocheng and Zhou, Sean, Credit Rating Prediction Through Supply Chains: A Machine Learning Approach (October 27, 2021). Production and Operations Management 31(4), 1613-1629, Available at SSRN: https://ssrn.com/abstract=3951507
[4] Berg, T, V Burg, A Gombović and M Puri (2018): "On the rise of fintechs-credit scoring using digital footprints", NBER Working Papers, 24551, April.
[5] Matz SC, Menges JI, Stillwell DJ, Schwartz HA. Predicting individual-level income from Facebook profiles. PLoS One. 2019 Mar 28;14(3): e0214369. doi: 10.1371/journal.pone.0214369. PMID: 30921389; PMCID: PMC6438464.
[6] Liu, Xinghai and Ding, Wei and Jia, Hongyu. Big Data's Application and Inspiration on Credit Scoring --- Taking U.S. Fintech Company Zest Finance as An Example(2014). China Academic Journal Publishing House. DOI:10.19409/j.cnki.thf-review.2014.10.024
[7] Fuster, A, P Goldsmith-Pinkham, T Ramadorai and A Walther: The effect of machine learning on credit markets(2019), VoxEU, 11 January 2019.
[8] Jessica J. Gonzalez, Communications Technology Use in Communities of Color: A Presentation to the FCC Technology Transitions Task Force, NATIONAL HISPANIC MEDIA COALITION, (Mar. 18, 2013), https://transition.fcc.gov/technologies_transitions_policy_task_force/Panel_2-Jessica_Gonzalez.pdf [https://perma.cc/U9PZ-LMLY].
[9] Godbillon-Camus, Brigitte and Godlewski, Christophe J., Credit Risk Management in Banks: Hard Information, Soft Information and Manipulation (December 2005). Available at SSRN: https://ssrn.com/abstract=882027
[10] Matz SC, Menges JI, Stillwell DJ, Schwartz HA. Predicting individual-level income from Facebook profiles. PLoS One. 2019 Mar 28;14(3): e0214369. doi: 10.1371/journal.pone.0214369. PMID: 30921389; PMCID: PMC6438464.